THE RESULT

"Promoted at 3x their peers"? Against real peers, 41% against 38%.

A Director of Leadership Development at a 40,000-person energy equipment maker checks her own programme's claims before the CHRO takes them to the talent review, with any AI chat her company allows.

THE RESULT What Amara sends the CHRO for the talent review. Fictional company.

HALVERN POWER SYSTEMS · TALENT REVIEW · DEC 8

Emerging Leaders: what the evidence shows

For: Henrik Dahl, CHRO
From: Amara Osei

THE SHORT ANSWER

Graduates are promoted at about the same rate as other high-potential managers: 41% (19 of 46) against 38% (34 of 90). We can't yet say the programme makes the difference, and we're setting up a fair test. · Input 3

THE CLAIMS

3x their peers **Wrong peers** Compared with all managers	**2024 cohort** **Window open** 18 months end December 2026	**Asia 83%** **5 of 6 people** Too few to judge	**360 rise** **12 of 46 replied** Not the same people

HOW WE'LL FIND OUT

2027 COHORT — **A fair staggered start**
Equally eligible managers are assigned to 2027 or 2028 starts by a fair draw. We compare early development first, then promotion rates once both groups have had the same follow-up time. · Input 3

ALL COHORTS — **Count everyone selected, and pair the 360s**
Outcomes for all 58, including the 12 who left or withdrew. Before-and-after 360s from the same people. · Inputs 3, 4

CAUGHT ON REVIEW

1. "3x their peers" **compares high potentials with all managers**
2. Against other high potentials, it's **41% against 38%**
3. The 41% **leaves out the 12 who didn't graduate**

UC-001 · SC-14 · REV B	
FOR **Director, Leadership Development**	SECTOR **Energy equipment**
COMPANY **40,000 people · 25 sites · 15 countries**	TASK **Show the CHRO whether the leadership programme works**
WRITTEN FOR **ChatGPT · Claude · Gemini · Copilot**	
AI TEST **Claude, Nov 13, 2026 · 6 of 6 found**	

STEP 0 · THE MESS

What landed on Amara's desk

A CHRO who wants to say the programme works, a proud summary from her own team, and the HR data underneath. The data includes the comparison the summary didn't use.

EMAIL · CHRO [1]

For the talent review

"I'd like to tell the talent review that Emerging Leaders works."

1 **A conclusion before the evidence.**

PROGRAMME SUMMARY [2]

Results

"Promoted at 3x the rate of their peers! … Asia: 83%."

2 **Which peers?**

HR COHORT DATA [3]

2023 and 2024

58 selected, 46 graduated. High-potential managers not in the programme: 34 of 90 promoted.

3 **The fair comparison.**

360 SUMMARY [4]

Before and after

3.6 before, 4.1 after. Before: all 46. After: 12 completed it.

4 **Two different groups.**

YOU'RE NOT ALONE

40%

of leaders rate leadership quality in their organisation as high.

DDI Global Leadership Forecast 2025 (10,796 leaders and 2,185 HR professionals in more than 50 countries). Source

STEPS 1–3 · HOW IT WORKS

Three steps, one chat

Here the AI checks whether a programme deserves the credit it claims. The prompt asks who is in each group.

1 Open a new chat

Use the AI tool your company approves: ChatGPT, Claude, Gemini or Copilot.

2 Paste the prompt and the data

Copy the prompt and the four fictional inputs from the appendix into one message.

3 Decide how you'll find out

Check the key comparisons yourself, then choose one fair test you'll run before next year's review.

THE PROMPT, PART BY PART

COPY-READY VERSION IN THE APPENDIX

I'm Director of Leadership Development at an energy equipment maker with about 40,000 people in 15 countries. Our CHRO wants one page on whether our Emerging Leaders Program works, for the annual talent review on December 8. Today is November 23. Below are his email, the programme team's summary, HR's cohort data and a 360 feedback summary.

1 Who you are, who wants the page, for which meeting, and today's date. The date lets it check time windows.

Please give me:

1. **Does the evidence show the programme works?** Two sentences.
2. **Each claim in the programme team's summary**: what it compares, and whether it holds.
3. **What we can honestly tell the talent review.**
4. **What evidence we'd need** to know whether it works.

2 A short answer, each claim tested, what you can honestly say, and what evidence you'd need.

Check who is in each group being compared, who is missing from the numbers, whether each time window is complete, and how many people each figure rests on. Don't add facts that aren't in the inputs.

3 The heavy lifting: who is in each group, who is missing, whether windows are complete, and how many people each figure rests on. It found all six problems.

WHERE TO PASTE

Every tool has one message box. Paste the prompt, add a blank line, paste the data, send. Exports can be attached as files if your tool allows it.

SAFE TO TRY

Practise with the fictional data first. Real talent and promotion data is sensitive: use only your company's approved tool.

THE AI'S FIRST ANSWER

What Claude actually wrote

Exactly as it came back, not edited. Read it the way Amara did: check each point against the source it cites.

UNEDITED Claude · fresh session, no other context · Nov 13, 2026 · one run

1. Does the evidence show the programme works?

No. The evidence doesn't show that it works or that it doesn't. Measured against the right peer group, graduates were promoted at 41% (19 of 46) and other high-potential managers at 38% (34 of 90). That gap is small, it leaves out the 12 people who didn't graduate, and it includes a cohort whose 18 months aren't over yet.

2. Each claim in the programme team's summary

Claim	What it compares	Does it hold?
"Graduates are promoted at 3x the rate of their peers"	41% for graduates against 14% for "all managers" (41 ÷ 14 ≈ 2.9)	**No.** "All managers" aren't the graduates' peers. Graduates were picked because they were rated high potential. The inputs give the right comparison: high-potential managers not in the programme, 34 of 90 promoted (38%). Against that group graduates are at about 1.1x, not 3x. The 14% figure also comes with no headcount, no source and no time window. `found`
"41% of graduates were promoted within 18 months"	19 promoted out of 46 graduates (11 + 8)	**The arithmetic is right (41.3%), but the figure overstates.** (a) It counts graduates only. 58 were selected, and the 12 who didn't graduate (7 left the company, 5 withdrew) are left out. Counting everyone selected, at least 19 of 58 (33%) were promoted, which is below the 38% for non-participants. Whether any of the 5 who withdrew were promoted is unknown. (b) The 2024 cohort graduated in June 2025, so its 18-month window doesn't close until December 2026. On November 23 it has about 17 of 18 months, assuming the current year is 2026. Its 8 of 22 is therefore an incomplete count. The 2023 cohort's window is complete: 11 of 24 (46%). (c) For the 34 of 90 comparison, the inputs don't say whether every member has had a full 18 months, or whether it includes people who left. `A: like for like`
"Results in Asia were exceptional: 83% promoted"	5 of 6 Asia graduates, both cohorts	**No.** It rests on 6 people, so one promotion more or less moves it by about 17 points. It has the same problems as the overall figure: graduates only, partly incomplete window. And there's no Asia comparison group. `found`
360 scores rose from 3.6 to 4.1 (from the 360 summary, not the team's summary, but likely to be cited)	The average of all 46 graduates before, against 12 graduates after	**No.** These are different groups. Only 12 of 46 (26%) completed the "after" 360, and we don't know who they were or what they scored before. If the people who chose to complete it were the ones doing well, the rise could come from that alone. `found`

3. What we can honestly tell the talent review

- 58 high-potential managers were selected across two cohorts. 46 graduated, 7 left the company and 5 withdrew.
- 19 of the 46 graduates (41%) have been promoted within 18 months. For the 2023 cohort, whose window is complete, it's 11 of 24 (46%). The 2024 cohort's window is still open until December 2026. `found`
- High-potential managers who weren't in the programme were promoted at 38% (34 of 90) over the same period. That is similar to the graduates' rate. Counting everyone selected, including those who left or withdrew, the programme group's rate is at least 33%. `found`
- So today we can't say the programme raises promotion rates. Any difference we see is small and could come from how people are selected and who completes the programme.
- The 360 improvement can't be relied on, because only 12 of 46 graduates completed the follow-up.
- The Asia figure (5 of 6) is too small to draw a conclusion from.
- Suggested line for Henrik: "Participants are being promoted at roughly the same rate as comparable high-potential managers. We don't yet have evidence that the programme itself makes the difference, and here is what we're doing to find out." `B: say what's next`

4. What evidence we'd need to know whether it works

- **A like-for-like comparison group:** high-potential managers with similar ratings, level, region and tenure who weren't in the programme, tracked over the same complete 18-month windows. Better still, a waitlist or staggered start, so that selection doesn't decide who takes part. `C: pick one test`
- **Everyone who was selected, counted:** outcomes for all 58, including the 7 who left and the 5 who withdrew, and the same treatment of leavers in the comparison group.
- **Complete windows only:** report the 2024 cohort after December 2026, and confirm the 34 of 90 figure uses complete windows.
- **A sourced, defined 14% figure** if it's used at all: who counts as a manager, which period, how many people.

- **Paired 360 data:** before and after scores for the same people, with a high completion rate, and ideally the same 360 for the comparison group.
- **Outcomes beyond promotion,** such as retention, performance ratings and later performance in the promoted role. Promotion can also reflect how visible the programme makes people.
- **Enough people for regional cuts:** with 6 people in Asia, pool more cohorts before reporting by region.
- **Timing:** Henrik asked for the page by December 1, a week before the December 8 review. Because the 2024 window closes in December, a fuller update could follow after the review.

`found` caught a planted problem `A–C` Amara changes it; see the next section

THE HUMAN PART

What Amara changed, and why

Claude found every planted problem. Amara removed one comparison that mixes two differently counted groups, and turned "we don't know" into a plan to find out.

WHAT CLAUDE FOUND: 6 OF 6 PLANTED PROBLEMS

One run

- ✓ "3x" uses the wrong peers
- ✓ Real peers 41% against 38%
- ✓ Leaves out 12 who didn't graduate
- ✓ 2024 window still open
- ✓ Asia is 5 of 6 people
- ✓ 360 after-score from 12 of 46

THREE EDITS, ONE CATCH AND TWO JUDGMENT CALLS

	CLAUDE WROTE	AMARA'S VERSION	WHY · ACCURACY
A	"At least 19 of 58 (33%) were promoted, which is below the 38% for non-participants."	**33% stays as context. The page compares 41% with 38% only, and says the comparison group's counting is unconfirmed.**	The 58 include people who left; the AI itself says we don't know whether the 90 do. Calling one lower than the other compares two groups counted differently.

	CLAUDE WROTE	AMARA'S VERSION	WHY · AUDIENCE
B	"We don't yet have evidence that the programme itself makes the difference, and here is what we're doing to find out."	**"About the same rate as other high potentials. We can't yet say the programme makes the difference, and we're setting up a fair test."**	Henrik will repeat this line in the room. It needs the number and a plan, so it lands as rigour, not as bad news.

	CLAUDE WROTE	AMARA'S VERSION	WHY · DECISION
C	Seven kinds of evidence, including "Better still, a waitlist or staggered start".	**One test: equally eligible managers assigned to 2027 or 2028 starts by a fair draw, plus counting everyone selected and pairing the 360s.**	Seven ideas become none. Amara picked one, and set it up fairly: a staggered start only works if who starts first is drawn, not chosen. Promotion results come later, once both groups have had equal time.

What only you can do Tell your own CHRO that your own programme's headline doesn't hold, and replace it with a test you'll actually run. The AI found the right peers. You decide how to find out.

YOUR TURN

Make it yours

Use it whenever a programme takes credit for an outcome: promotions, retention, performance, safety.

YOUR PROMPT: FILL IN THE BRACKETS

I'm the **[your role]** at a **[industry]** company with about **[number]** people. **[Who]** wants **[a page or claim]** on whether **[the programme]** works, for **[the meeting]** on **[date]**. Today is **[date]**. Below are **[the request, the programme's own claims, the cohort data and any survey data]**.

Please give me: **1. Does the evidence show the programme works?** Two sentences. **2. Each claim** in **[the summary]**: what it compares, and whether it holds. **3. What we can honestly tell [the audience]**. **4. What evidence we'd need** to know whether it works.

Check who is in each group being compared, who is missing from the numbers, whether each time window is complete, and how many people each figure rests on. Don't add facts that aren't in the inputs.

FIVE CHECKS BEFORE YOU SEND

1. The comparison group was selected the same way as the participants.
2. People who left or withdrew are counted, or named as missing.
3. Every time window is complete, or flagged.
4. No conclusion rests on a handful of people.
5. Before and after figures come from the same people.

MAKE IT REPEATABLE (OPTIONAL)

Save the prompt once and reuse it for any programme review: a custom GPT in ChatGPT, a project in Claude, a Gem in Gemini, or an agent in Copilot. Some companies switch these off; copy and paste always works.

WITH REAL DATA

Talent ratings and promotion data are sensitive personal data. Use only the AI tool your company approves, and follow its rules.

WHAT THIS EXAMPLE SHOWS

One run of this prompt in Claude found all six problems planted in fictional programme data and invented nothing.

WHAT IT DOESN'T SHOW

That every tool, setting or run will do the same, how much time it saves you, or whether any real programme works. Halvern Power Systems and its people are fictional.

APPENDIX

Copy-ready: the prompt and the four inputs

Copy the prompt, then all four inputs, into one message in a new chat. Everything here is fictional and safe to paste.

THE PROMPT

I'm Director of Leadership Development at an energy equipment maker with about 40,000 people in 15 countries. Our CHRO wants one page on whether our Emerging Leaders Program works, for the annual talent review on December 8. Today is November 23. Below are his email, the programme team's summary, HR's cohort data and a 360 feedback summary.

Please give me:

1. **Does the evidence show the programme works?** Two sentences.
2. **Each claim in the programme team's summary**: what it compares, and whether it holds.
3. **What we can honestly tell the talent review.**
4. **What evidence we'd need** to know whether it works.

Check who is in each group being compared, who is missing from the numbers, whether each time window is complete, and how many people each figure rests on. Don't add facts that aren't in the inputs.

[1] EMAIL FROM HENRIK DAHL, CHRO

Subject: Emerging Leaders, for the talent review

Amara, I'd like to tell the talent review that Emerging Leaders works. Can you give me one page on the evidence by December 1?

[2] SUMMARY FROM THE PROGRAMME TEAM

Graduates are promoted at 3x the rate of their peers! 41% of graduates were promoted within 18 months, against 14% of all managers. Results in Asia were exceptional: 83% promoted.

[3] COHORT DATA FROM HR

Cohort	Selected	Graduated	Did not graduate	Graduates promoted within 18 months
2023 (graduated June 2024)	30	24	6 (4 left the company, 2 withdrew)	11
2024 (graduated June 2025)	28	22	6 (3 left the company, 3 withdrew)	8

Who is selected: managers rated high potential in the annual talent review. High-potential managers not in the programme, same period: 34 of 90 promoted within 18 months. Asia, both cohorts: 5 of 6 graduates promoted.

[4] 360 FEEDBACK SUMMARY

Average 360 score: 3.6 before the programme, 4.1 after (out of 5). Before: all 46 graduates. After: 12 graduates completed it.